

57672

From: Davis, Natalie  
Sent: Monday, January 07, 2002 11:12 AM  
To: STIC-Biotech/ChemLib  
Subject: 09/823101

Please search SEQ ID NO: 1-3

Natalie A. Davis, PhD  
Patent Examiner  
Art Unit 1642  
CM1, Rm 8B13  
Mailbox 8E12  
Ph (703) 308-6410

CRFE

Point of Contact:  
Mona Smith  
Technical Info. Specialist  
CM1 12C14 Tel: 308-3278

Searcher: M. Smith  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 1/7/02  
Date Completed: 1/6  
Searcher Prep/Review: 10  
Clerical: \_\_\_\_\_  
Online time: 10

TYPE OF SEARCH:  
NA Sequences: 3  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST(where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2002, 08:44:55 ; Search time 3002.62 seconds  
(without alignments)  
841.018 Million cell updates/sec

Title: US-09-823-101-2

Perfect score: 235

Sequence: 1 acctctaccactgctgac.....ttcttagcaagtgaagctt 235

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estfun:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estom:\*

5: em\_estpl:\*

6: em\_estro:\*

7: em\_estro:\*

8: em\_estov:\*

9: em\_hic:\*

10: gb\_est1:\*

11: gb\_est2:\*

12: gb\_hic:\*

13: gb\_gss:\*

14: em\_gss\_fun:\*

15: em\_gss\_hum:\*

16: em\_gss\_inv:\*

17: em\_gss\_pin:\*

18: em\_gss\_pro:\*

19: em\_gss\_rod:\*

20: em\_gss\_vrt:\*

21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	49.4	21.0	511	13	AZ761057 1M0555K12
2	39.8	16.9	432	13	BH078711 RPCI-24-3
3	38.6	16.4	578	13	AZ376465 1M0130P16
4	38.6	16.4	591	13	AZ375783 1M0129P16
5	38.2	16.3	606	13	BH060091 RPCI-24-3
6	38.2	16.2	417	10	BE253748 601112353
7	37.8	16.1	469	13	AQ655258 Sheared D
8	37.8	16.1	651	13	AQ643656 RPCI93-Dp
9	37.8	16.1	758	11	BG53874 102403800
10	37.6	16.0	863	13	CNS012SP
C 11	37.6	16.0	1101	13	AL101971 Drosophil
12	37.4	15.9	522	11	AL067324 Drosophil
					BG284982 602409245

C 13	36.8	15.7	560	13	AZ429688
14	36.4	15.5	422	10	AI653440
C 15	36.4	15.5	496	13	AZ128979 OSJNB006
16	36.2	15.4	409	11	H60465
C 17	36.2	15.4	841	10	AL569885
18	36.2	15.4	845	11	BF578564
C 19	36.2	15.4	923	13	CNS04NPT
20	35.8	15.2	1101	13	CNS002B1
C 21	35.6	15.1	1340	11	BE890770
22	35.4	15.1	786	13	CNS02LRS
C 23	35.4	15.1	997	13	CNS005TE
24	35.4	15.1	1101	13	CNS006C9
C 25	35.2	15.0	439	10	AI879059
26	35.2	15.0	482	10	AW403656
C 27	35.2	15.0	622	10	AW953270
28	35.2	15.0	629	11	BG684818
C 29	35.2	15.0	780	11	BF128982
30	35.2	15.0	835	11	BF337210
C 31	35.2	15.0	424	13	AZ496846
32	34.8	14.8	291	10	BB528060
C 33	34.8	14.8	501	13	AZ193303
34	34.8	14.8	873	13	AZ128128
C 35	34.8	14.8	984	11	BG497837
36	34.6	14.7	493	10	AI399104
C 37	34.6	14.7	690	11	BI197022
38	34.6	14.7	783	11	BG854194
C 39	34.6	14.7	813	11	BF684008
40	34.6	14.7	1100	13	CNS05P6E
C 41	34.4	14.6	609	13	AQ717767
42	34.4	14.6	879	11	BG778671
C 43	34.4	14.6	987	13	CNS00418
44	34.4	14.6	1652	11	BF863487
C 45	34.2	14.6	484	13	AQ198109

ALIGNMENTS

RESULT 1  
AZ761057/c  
LOCUS  
DEFINITION  
1M0555K12F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0555K12 F, DNA sequence.  
ACCESSION  
AZ761057.1  
VERSION  
GSS.  
KEYWORDS  
house mouse.  
SOURCE  
Mus musculus  
ORGANISM  
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 511)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rellly  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhauser,A.  
and Wright,D.,Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
JOURNAL  
Unpublished (2000)  
COMMENT  
Contact: Robert B. Weiss  
University of Utah  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0555 row: K column: 12  
Seq primer: CGTTGTAACGACGCCAGT  
Class: plasmid ends  
High quality sequence stop: 511.  
Location/Qualifiers  
1. .511



us-09-823-101-2.rst

Tue Jan '8 10:26:12 2002

```

source
1. .578
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0130P16"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWB42 (g114732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 156 a 119 c 160 g 143 t
ORIGIN

Query Match 16.4%; Score 38.6; DB 13; Length 578;
Best Local Similarity 56.8%; Pred. No. 3.1;
Matches 71; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY . 104 aaggaggatccaaacccctggagaagacagcagtagagtgaggcgagtgaggagcag 163
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 396 AAGGAGCGAACCAATCCCTATGATGAATGAAGGCCAATCACTGGCGAGTAGGATCGGA 455

QY . 164 gaaaggttggtgacagatcaggaggtgtctgacctttttcttgagaaattctagg 223
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 456 GAGAGGAAGAGAGAGGGGGGAGAGAGAGAGAGGGCCCTTTTGGATGAGACAGATGGAG 515

QY 224 caagt 228
||| |
Db 516 CAACT 520

RESULT 4
AZ375783 591 bp DNA GSS 02-OCT-2000
LOCUS IM0129P16F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION
ACCESSION AZ375783
VERSION AZ375783.1 GI:10489483
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 591)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA

TITLE
JOURNAL
COMMENT

1. .591
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0129P16"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWB42 (g114732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 158 a 124 c 164 g 145 t
ORIGIN

Query Match 16.4%; Score 38.6; DB 13; Length 591;
Best Local Similarity 56.8%; Pred. No. 3.1;
Matches 71; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 104 aaggggagatccaaacccctggagaagacagcagtagaagtgaggcgagtgaggagcag 163
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 397 AAGGAGCGAACCAATCCCTATGATGAATGAAGGCCAATCACTGGCGAGTAGGATCGGA 456

QY 164 gaaaggttggtgacagatcaggaggtgtctgacctttttcttgagaaattctagg 223
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 457 GAGAGGAAGAGAGAGGGGGGAGAGAGAGAGAGAGGGCCCTTTTGGATGAGACAGATGGAG 516

QY 224 caagt 228
||| |
Db 517 CAACT 521

RESULT 5
BH060091 606 bp DNA GSS 18-JUL-2001
LOCUS RPCI-24-326K19.TV RPCI-24 Mus musculus genomic clone RPCI-24-326K19
DEFINITION
ACCESSION BH060091
VERSION BH060091.1 GI:14870888
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 606)
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.

```

# TITLE JOURNAL COMMENT

Mouse BAC End Sequences from Library RPCI-24  
Unpublished (1999)  
Other\_GSSs: RPCI-24-326K19.TJ  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-24. For BAC  
library availability, please contact Pieter de Jong  
(pdejong@mail.cho.org). Clones may be purchased from BACPAC  
Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end  
plate: [http://www.tigr.org/tldb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html)  
Plate: 326 row: K column: 19  
Seq primer: T7  
Class: BAC ends.

## FEATURES source

Location/Qualifiers  
1. 606  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-24-326K19"  
/clone\_lib="RPCI-24"  
/sex="Male"  
/cell\_type="Spleen/Brain"  
/note="Vector: pTANBAC1; Site\_1: BamHI; Site\_2: BamHI;  
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The  
library was cloned in the pTANBAC1 cloning vector at the  
BamHI sites using MboI partially digested male C57BL/6J  
DNA."

BASE COUNT 162 a 130 c 174 g 140 t

Query Match 16.3%; Score 38.2; DB 13; Length 606;  
Best Local Similarity 56.9%; Pred. No. 4;  
Matches 70; Conservative 0; Mismatches 53; Indels 0; Gaps 0;  
QY 104 aaggaggatccaaacctggaagacacagcagtagaagtgagcagtgaggagcag 163  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 396 AAGGAGGAGCAACCAATCCCTATGATATAAGGAGCCATCATCTGGCGAGTAGGTCGGA 455  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 164 gaaaggttggtgacagatcaggagggtgtctgacccctttcttgaggaaattctagg 223  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 456 GAGGGAAGAGAGAGGGGGGAGAGAGAGAGAGAGGCGCCCTTTTGGATGAGGACAGATGCAG 515  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 224 caa 226  
|||  
Db 516 CAA 518

RESULT 6  
BE253748  
LOCUS  
DEFINITION BE253748 417 bp mRNA EST 13-JUL-2000  
601112353F1 NIH\_MGC\_16 Homo sapiens cDNA clone IMAGE:3353172 5',  
mRNA sequence.  
ACCESSION BE253748  
VERSION BE253748.1 GI:9124168  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS 1 (bases 1 to 417)  
TITLE NIH-MGC <http://imgc.ncbi.nih.gov/>.  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at: [image.llnl.gov](http://image.llnl.gov)  
Plate: LCM153 row: i column: 13  
High quality sequence stop: 410.

## FEATURES source

Location/Qualifiers  
1. 417  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3353172"  
/clone\_lib="NIH\_MGC\_16"  
/tissue\_type="retinoblastoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: eye; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5',  
adaptor: GGCACGAG(G). Library constructed by Ling Hong  
in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."  
BASE COUNT 110 a 103 c 136 g 68 t  
ORIGIN

Query Match 16.2%; Score 38; DB 10; Length 417;  
Best Local Similarity 51.1%; Pred. No. 4.2;  
Matches 89; Conservative 0; Mismatches 85; Indels 0; Gaps 0;  
QY 41 agagagtgaggagatgaagtcaggtctctcatagcaaatctatgcttacaatcccg 100  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 223 AAGAGCGAGAAATGGAAGAGAGCGCCCTGCTAACGGGAATGCTAATGAGGAATGGG 282  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 101 agaaaggggagatccaaacctggaagacacagcagtagaagtgaggcagtgaggagg 160  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 283 AGCAGGAGCTGCACATGAGGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 342  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 161 caggaaaggttggtgacagatcaggagggtgtctgacccctttcttgaggaa 214  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 343 AGAAGAAGATGGTGTGAGTGGGTGCAGGAGAGAGAGATGAGATGAGGAA 396  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7  
AQ655258  
LOCUS  
DEFINITION AQ655258 469 bp DNA GSS 22-JUN-1999  
Sheared DNA-20M16.TF Sheared DNA Trypanosoma brucei genomic clone  
Sheared DNA-20M16, DNA sequence.  
ACCESSION AQ655258  
VERSION AQ655258.1 GI:5148444  
KEYWORDS GSS.  
SOURCE Trypanosoma brucei.  
ORGANISM Trypanosoma brucei.  
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
Trypanosoma.  
1 (bases 1 to 469)  
AUTHORS El-Sayed, N., Zhao, S., Gill, S., Suh, E., Malek, J., Fujii, C.,  
Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J.,  
Fraser, C. and Adams, M.  
TITLE Determination of clone end sequences from Trypanosoma brucei GUTat  
10.1 sheared DNA library  
JOURNAL Unpublished (1999)  
COMMENT Other\_GSSs: Sheared DNA-20M16.TR  
Contact: Najib M. El-Sayed  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: nelsayed@tigr.org  
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared  
DNA library constructed at TIGR. Clones will be available for  
distribution through ATCC. Sheared DNA end sequences search page:







RESULT 13  
AZ429688/C  
LOCUS

```

Qy 163 ggaaggttggtgacagatcagg 186
||||| ||| ||| ||| ||| |||
Db 362 GGAAGGTGAGAGGGAGAGAGG 339

RESULT 14
LOCUS AI653440
DEFINITION tq94D07.x1 NCI_CGAP_Ov23 Homo sapiens cDNA clone IMAGE:2216437 3'
similar to contains MER22.t3 MSRI repetitive element ;, mRNA
sequence.
ACCESSION AI653440
VERSION AI653440
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 422)
AUTHORS NCI-CGAP
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 1609 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 412.
FEATURES
source
1..422
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2216437"
/clone_lib="NCI_CGAP_Ov23"
/tissue_type="tumor, 5 pooled (see description)"
/lab_host="DH10B"
/notes="Organ: ovary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.35 Kb. Tumor types include: mixed
Mullerian tumor, papillary serous, clear cell, spindle
cell. All are primary tumors, metastasis positive. Life
Technologies catalog #: 11534-013"
BASE COUNT 112 a 71 c 156 g 72 t 11 others
ORIGIN

Query Match 15.5%; Score 36.4; DB 10; Length 422;
Best Local Similarity 53.3%; Pred. No. 11;
Matches 64; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

Qy 105 agggagatccaaacctgaagaagacagcgtagtgagtgaggcgaggcagg 164
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 244 ATGGGGAGACGACAGAGAGGGGGGACAGAGGGGGCGCGGTTCGTCGAGG 303

Qy 165 aaagttgggtgacagatcagggaggtgtctgacaccttttcttgaggaattcttaggc 224
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 304 AGAGGGAGGGAGGGGAGAGGGGGGACNNNNCTTNTNCNTNATCAACAGGC 363

RESULT 15
LOCUS AZ128979/c
DEFINITION OSJNB0068N24f CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
clone OSJNB0068N24f, DNA sequence.

```

```

ACCESSION AZ128979
VERSION AZ128979.1
KEYWORDS GI:8206794
SOURCE GSS.
ORGANISM Oryza sativa.
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 496)
AUTHORS Wing, R.A. and Dean, R.A.
TITLE A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL Unpublished (1998)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: GTAAACGACGCGCAGTG
Class: BAC ends
High quality sequence start: 37
High quality sequence stop: 161.
FEATURES
source
1..496
/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="OSJNB0068N24f"
/clone_lib="CUGI Rice BAC Library (EcoRI)"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/notes="Vector: pBACIndigo; Site_1: EcoRI; Site_2: EcoRI;
Rice is the most important food crop in the world. Half of
the world population, especially those inhabiting highly
populated areas of the humid tropics and subtropics, rely
on rice as their primary source of carbohydrate.
Monocotyledonous rice is a diploid plant (2n=24) with a
haploid genome equivalent of 431 Mbp (Arumuganathan and
Earle, 1991). The relatively small genome of rice, three
times larger than that of Arabidopsis, makes it suitable
for genomic studies. In order to facilitate positional
cloning, physical mapping and genome sequencing of rice,
we have constructed a BAC library from Oryza sativa,
Nipponbare variety using EcoRI as the cloning enzyme. The
library contains 55,296 clones with an average insert size
of 121 Kb providing approximately 15 haploid genome
equivalents. The deep coverage allows the isolation a
particular sequence with a probability of 99.9 %. Three
high density filters, each containing 18,432 clones
(doubly spotted), represent the whole library for colony
screening and can be requested from the Clemson University
BAC/EST Resource Center (www.genome.clemson.edu)."
BASE COUNT 86 a 232 c 46 g 130 t 2 others
ORIGIN

Query Match 15.5%; Score 36.4; DB 13; Length 496;
Best Local Similarity 49.5%; Pred. No. 12;
Matches 91; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Qy 23 ggtgagtggtgtcccaaggagtagagagataagagtcaggtctcatagccaata 82
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 372 GGAGGGGAGGGGNGGAGAGAGAGGGAGGTGNGGGGGGAGGAGGGGGAAG 313

Qy 83 ctatggttcagtcctccgaggaagggtgagatccaaaccttggaagaagacagcagtag 142
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 312 GGGAGGGGGGAAGGGGGAGGGAGGGGGGAGAGGGGGGGGGAGGAGGGGAGG 253

Qy 143 aagtggcgagtgaggagcaggaaaggttggtgacagatcaggaggggtgtctgacctt 202
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 252 GAGGGGAGGGGAGGGGAGGAGGAGGGGGGAGGAGGAGGAGGGGAGGGGTGACCCC 193

```

Qy 203 tttc 206  
|||  
Db 192 ATTC 189

Search completed: January 8, 2002, 08:45:00  
Job time: 5607 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2002, 08:47:16 ; Search time 116.06 Seconds  
(without alignments)  
458.576 Million cell updates/sec

Title: US-09-823-101-2  
Perfect score: 235  
Sequence: 1 acctctaccactgctgtac.....ttcttagcgaagtgaagctt 235

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues  
Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	43.8	18.6	7218	1	US-08-232-463-14
C 2	34.8	14.8	3482	3	US-09-111-085-3
C 3	34	14.5	2184	4	US-09-307-265A-2
C 4	31.8	13.5	289	4	US-09-007-005-17
C 5	31.8	13.5	289	4	US-09-244-796-17
C 6	31.4	13.4	2061	2	US-08-835-170-1
C 7	31.4	13.4	2061	4	US-09-359-257-1
C 8	31.4	13.4	2327	2	US-08-835-170-3
C 9	31.4	13.4	2327	4	US-09-359-257-3
C 10	31	13.2	6171	1	US-08-459-568-1
C 11	31	13.2	6171	2	US-08-399-411-1
C 12	31	13.2	6171	3	US-08-516-859A-1
C 13	30.8	13.1	2188	1	US-07-865-662F-10
C 14	30.8	13.1	2188	4	US-08-374-219B-10
C 15	30.2	12.9	1461	1	US-08-587-670A-1
C 16	30.2	12.9	1461	4	US-09-061-674-1
C 17	30.2	12.9	80246	4	US-09-078-294-4
C 18	30.2	12.9	80595	4	US-09-078-294-3
C 19	29.8	12.7	2580	3	US-09-050-863-2
C 20	29.8	12.7	5452	2	US-09-130-114-1
C 21	29.8	12.7	9600	4	US-08-910-114-1
C 22	29.8	12.7	10596	1	US-07-884-811-15
C 23	29.8	12.7	10596	1	US-07-885-971-15
C 24	29.8	12.7	10596	1	US-08-087-783A-15
C 25	29.8	12.7	10596	1	US-08-194-088B-15
C 26	29.8	12.7	10596	2	US-08-194-087-15
C 27	29.8	12.7	10596	5	PCT-US93-04648-15

C 28	29.6	12.6	1200	3	US-08-617-860B-4	Sequence 4, Appl1
C 29	29.6	12.6	1931	2	US-09-130-114-2	Sequence 2, Appl1
C 30	29.4	12.5	19056	4	US-09-272-032-8	Sequence 8, Appl1
C 31	29.2	12.4	197	1	US-08-392-678-5	Sequence 5, Appl1
C 32	29.2	12.4	197	1	US-08-457-304A-5	Sequence 5, Appl1
C 33	29.2	12.4	197	1	US-08-456-701A-5	Sequence 2, Appl1
C 34	29	12.3	5361	4	US-08-973-462-2	Sequence 1, Appl1
C 35	29	12.3	6152	4	US-08-973-462-1	Sequence 2, Appl1
C 36	28.6	12.2	1540	2	US-08-943-600A-2	Sequence 2, Appl1
C 37	28.4	12.1	325	2	US-08-332-766A-11	Sequence 11, Appl1
C 38	28.4	12.1	975	6	5340934-10	Patent No. 5340934
C 39	28.4	12.1	2089	1	US-08-552-142A-1	Sequence 1, Appl1
C 40	28.4	12.1	2089	1	US-08-910-973-1	Sequence 1, Appl1
C 41	28.4	12.1	2089	5	PCT-US95-05741-1	Sequence 1, Appl1
C 42	28.4	12.1	35524	3	US-08-923-137-1	Sequence 1, Appl1
C 43	28.2	12.0	421	2	US-08-332-766A-25	Sequence 25, Appl1
C 44	28.2	12.0	1275	2	US-08-958-642-1	Sequence 1, Appl1
C 45	28.2	12.0	1275	3	US-08-778-394-3	Sequence 3, Appl1

ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-Fls  
US-08-232-463-14



```
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match      13.5%; Score 31.8; DB 4; Length 289;
Best Local Similarity 1.7%; Pred. No. 0.23;
Matches 3; Conservative 84; Mismatches 87; Indels 0; Gaps 0;

QY 17 tgacttggtgagtggtggtccaaaggagtaggagagataagagtcaggctctcatagc 76
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 24 utarururarcrarurarcrarurgrnrnrnrnrnrnrnrnrnrnrnrnrnrnr 83

QY 77 caaatactgcttcagtcgcagcaggaaggagagatccaaacctggagaagacaagg 136
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 84 nrsnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnr 143

QY 137 caqtagaagtggtgagtggtgagcaggaagggtggtgacagatcagggaag 190
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 144 nrsnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnr 197

RESULT 6
US-08-835-170-1
; Sequence 1, Application US/08835170
; Patent No. 5965420
; GENERAL INFORMATION:
; APPLICANT: Creasy, Caretha
; TITLE OF INVENTION: Human Protein Kinases HYAK3
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/835,170
; FILING DATE: 26-FEB-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Han, William T
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: GH50004
```

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5219
; TELEFAX: 610-270-4026
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2061 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-08-835-170-1

Query Match      13.4%; Score 31.4; DB 2; Length 2061;
Best Local Similarity 51.8%; Pred. No. 0.65;
Matches 71; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 93 cagtcctccgaggaaggagatccaaaccttggaagaagacagagcagtagaagtggcgca 152
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1514 CAGGCAGATGGGAGGCTTGCTTGTGGGGGTCGCTCAGGTAGGGTAAAAACGGGGT 1573

QY 153 gtggggagcaggaagggttgggtgacagatcagggagggtgtctgacctttttcttgagg 212
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1574 CCCCCAGGCAGCAAGACTGGGGGACAGCACTGAAAGGGTGTGATGACTACTTGTATTATA 1633

QY 213 aaattcttaggcaagtg 229
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1634 GAGTCTCTTGAAGAGGTG 1650

RESULT 7
US-09-359-257-1
; Sequence 1, Application US/09359257
; Patent No. 6165766
; GENERAL INFORMATION:
; APPLICANT: Xie, Wei
; APPLICANT: Creasy, Caretha
; TITLE OF INVENTION: HUMAN PROTEIN KINASES HYAK3
; FILE REFERENCE: GH50004XID1
; CURRENT APPLICATION NUMBER: US/09/359,257
; CURRENT FILING DATE: 1999-07-22
; EARLIER APPLICATION NUMBER: 08/835,170
; EARLIER FILING DATE: 1997-04-07
; EARLIER APPLICATION NUMBER: 60/040618
; EARLIER FILING DATE: 1997-03-05
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 2061
; TYPE: DNA
; ORGANISM: Human
US-09-359-257-1

Query Match      13.4%; Score 31.4; DB 4; Length 2061;
Best Local Similarity 51.8%; Pred. No. 0.65;
Matches 71; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 93 cagtcctccgaggaaggagatccaaaccttggaagaagacagagcagtagaagtggcgca 152
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1514 caggcagatgggaagggttctgttggtgggggtcgctcagctaggggttaaaaacgggggt 1573

QY 153 gtggggagcaggaagggttgggtgacagatcagggagggtgtctgacctttttcttgagg 212
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1574 cccccaggcagcaagactgggggacagcactgaaagggtgtgatgactactgtttata 1633

QY 213 aaattcttaggcaagtg 229
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1634 gagttcttgaagggtg 1650

RESULT 8
```





```
; LOCATION: 157..5275
US-08-459-568-1

Query Match      13.2%; Score 31; DB 1; Length 6171;
Best Local Similarity 62.0%; Pred. No. 1.3;
Matches 49; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 100 gaggaagggagatccaaaccctggaagacaagcagtagaagtggcgagtgagag 159
| ||| ||||| | | ||||| | | ||||| | | ||||| | | ||||| | | |||||
Db 3350 GGGAGGGGGAGGTGGAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3291

QY 160 gcaggaaggttgggtgac 178
| || | ||||| |||
Db 3290 GAAGAAAGTGTGGGGGAC 3272

RESULT 11
US-08-399-411-1/C
; Sequence 1, Application US/08399411
; Patent No. 5831008
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/399,411
; FILING DATE: 06-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1264
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6171 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 157..5275
; US-08-399-411-1

Query Match      13.2%; Score 31; DB 2; Length 6171;
Best Local Similarity 62.0%; Pred. No. 1.3;
Matches 49; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 100 gaggaagggagatccaaaccctggaagacaagcagtagaagtggcgagtgagag 159
| ||| ||||| | | ||||| | | ||||| | | ||||| | | ||||| | | |||||
Db 3350 GGGAGGGGGAGGTGGAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3291

QY 160 gcaggaaggttgggtgac 178
| || | ||||| |||
Db 3290 GAAGAAAGTGTGGGGGAC 3272

RESULT 12
US-08-516-859A-1/C
; Sequence 1, Application US/08516859A
; Patent No. 6069231
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/516,859A
; FILING DATE: 18-AUG-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/399,411
; FILING DATE: 06-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/292,683
; FILING DATE: 18-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1776
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6171 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 157..5275
; US-08-516-859A-1

Query Match      13.2%; Score 31; DB 3; Length 6171;
Best Local Similarity 62.0%; Pred. No. 1.3;
Matches 49; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 100 gaggaagggagatccaaaccctggaagacaagcagtagaagtggcgagtgagag 159
| ||| ||||| | | ||||| | | ||||| | | ||||| | | ||||| | | |||||
Db 3350 GGGAGGGGGAGGTGGAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3291

QY 160 gcaggaaggttgggtgac 178
| || | ||||| |||
Db 3290 GAAGAAAGTGTGGGGGAC 3272

RESULT 13
US-07-865-662F-10/c
; Sequence 10, Application US/07865662F
; Patent No. 5451670
; GENERAL INFORMATION:
; APPLICANT: Marcia M. Miller
; TITLE OF INVENTION: Restriction Fragment Length
; POLYMORPHISM TEST FOR HAPLOTYPE
; NUMBER OF SEQUENCES: 14
```

```

> STATE: D. C.
> COUNTRY: USA
> ZIP: 20004
> COMPUTER READABLE FORM:
> MEDIUM TYPE: Floppy disk
> COMPUTER: IBM PC compatible
> OPERATING SYSTEM: PC-DOS/MS-DOS
> SOFTWARE: Patent In Release #1.0, Version #1.30
> CURRENT APPLICATION DATA:
> APPLICATION NUMBER: US/08/374,219B
> FILING DATE: 18-JAN-1995
> CLASSIFICATION: 435
> PRIOR APPLICATION DATA:
> APPLICATION NUMBER: 07/865,662
> FILING DATE: 07-APR-1992
> APPLICATION NUMBER: 07/688,326
> FILING DATE: 22-APR-1991
> APPLICATION NUMBER: 07/588,922
> FILING DATE: 27-SEP-1990
> APPLICATION NUMBER: 07/413,301
> FILING DATE: 28-SEP-1989
> APPLICATION NUMBER: 07/210,405
> FILING DATE: 23-JUN-1988
> APPLICATION NUMBER: 07/130,529
> FILING DATE: 09-DEC-1987
> APPLICATION NUMBER: 07/068,176
> FILING DATE: 30-JUN-1987
> ATTORNEY/AGENT INFORMATION:
> NAME: Walker, Barbara W
> REGISTRATION NUMBER: 35,400
> REFERENCE/DOCKET NUMBER: 2124-113A
> TELECOMMUNICATION INFORMATION:
> TELEPHONE: 202 783-6040
> TELEFAX: 202 783-6031
> INFORMATION FOR SEQ ID NO: 10:
> SEQUENCE CHARACTERISTICS:
> LENGTH: 2188 base pairs
> TYPE: nucleic acid
> STRANDEDNESS: double
> TOPOLOGY: linear
> MOLECULE TYPE: DNA (genomic)
> US-08-374-219B-10

Query Match 13.1%; Score 30.8; DB 4; Length 2188;
Best Local Similarity 57.1%; Pred. No. 1;
Matches 56; Conservative 0; Mismatches 42; Indels 0; Gaps 0

Qy 123 ggaagaagacaagcagctagaagtggcgcgagtgggagcaggagaaaggtgtgggtgacagat 182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 145 GGACGAGACGACGAGATGAAGGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATGA 86

Qy 183 caggggaggtgtctgacctttttcttcttgaggaaattctt 220
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 85 TGATGATGATGAATAAGTCTCTTTTAGAGCAATTCTT 48

RESULT 15
US-08-587-670A-1/c
> Sequence 1, Application US/08587670A
> Patent No. 5824535
> GENERAL INFORMATION:
> APPLICANT: Kou, Guang-Hsiung
> APPLICANT: Wang, Chung-Hsiung
> APPLICANT: Lo, Chu-Fang
> TITLE OF INVENTION: IDENTIFICATION, PURIFICATION AND
> TITLE OF INVENTION: DETECTION OF WSBV (BACULOVIRUS ASSOCIATED WITH
> TITLE OF INVENTION: WHITE SPOT SYNDROME)
> NUMBER OF SEQUENCES: 14
> CORRESPONDENCE ADDRESS:
> ADDRESSEE: Fish & Richardson P.C.
> STREET: 225 Franklin Street
> CITY: Boston

```

```

; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/587,670A
; FILING DATE: 17-JAN-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsao, Y. Rocky
; REGISTRATION NUMBER: 34,053
; REFERENCE/DOCKET NUMBER: 07949/002001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1461 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-587-670A-1

```

```

Query Match      12.9%; Score 30.2; DB 1; Length 1461;
Best Local Similarity 56.6%; Pred. No. 1.4;
Matches 56; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 79 aatactatggttacagtcgccgaggaaggagagatccaaacccctggaagaagacaaggca 138
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 158 AATTTCATGTTGTCCTGAAGAAGATGACTTTGTATGGGAATCCATGGGAAGAAGCATGAGG 99
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 139 gtagaagtggcgagtgaggcagcaagaaggttggtga 177
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 98 TGAGAACCCAGGTGAGGGCGGCGGAGGAGGCTGGGTCA 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: January 8, 2002, 08:47:28  
Job time: 5680 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2002, 08:53:02 ; Search time 279.63 Seconds  
(without alignments)  
720.493 Million cell updates/sec

Title: US-09-823-101-2

Perfect score: 235

Sequence: 1 acctctaccactgctgac.....tctctagcgaagtgaagctt 235

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_1101.\*  
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	35.2	15.0	1099	21	Human ORFX ORF2753
C 2	34.8	14.8	1974	18	Porcine retrovirus
C 3	34.8	14.8	3482	20	Pig endogenous ret
C 4	34.8	14.8	8196	18	Porcine retrovirus
C 5	34.8	14.8	8209	18	Porcine retrovirus
C 6	34.6	14.7	500	21	Cat flea head and
C 7	34.4	14.6	626	22	Probe #12056 for g
C 8	34.4	14.6	626	22	Probe #16105 used
C 9	34.4	14.6	626	22	Probe #7814 used t
C 10	34.4	14.6	2784	18	Luciferase gene un
C 11	34	14.5	1954	22	Human polynucleoti

12	34	14.5	2182	22	AAI59351	Human polynucleoti
13	34	14.5	2184	22	AAD06295	Human Ras suppress
14	34	14.5	3251	22	AAH16802	Human CDNA sequenc
15	34	14.5	3253	22	AAH15878	Human CDNA sequenc
c 16	33.6	14.3	1297	21	AAA26368	Human secreted pro
17	33.2	14.1	575	22	AAI23315	Probe #13248 for g
18	33.2	14.1	575	22	AAI23315	Probe #17322 used
19	33.2	14.1	575	22	AAI08956	Probe #8947 used t
20	33.2	14.1	1969	22	AAI14105	Probe #4038 for ge
21	33.2	14.1	1969	22	AAI135486	Probe #4172 used t
22	33.2	14.1	1969	22	AAI03958	Probe #3949 used t
c 23	32.8	14.0	484	22	AAI38784	Human secreted pro
24	32.8	14.0	2043	19	AAV59728	Human protein enco
c 25	32.8	14.0	3149	22	AAH99797	Human Th2-specific
c 26	32.8	14.0	3868	21	AAAO9058	Human Th2-specific
c 27	32.8	14.0	4421	21	AACT6287	Human ORFX ORF1842
c 28	32	13.6	1535	22	AAZ93355	Sequence encoding
29	31.6	13.4	2800	22	AAD07131	Canine retinitis p
30	31.6	13.4	2803	22	AAD07132	Canine retinitis p
31	31.6	13.4	2805	22	AAD07130	Canine retinitis p
32	31.4	13.4	964	20	AAAX97739	Extended human sec
33	31.4	13.4	1221	21	AAAC37678	Arabidopsis thalia
34	31.4	13.4	1671	22	AAH99858	Human protein enco
35	31.4	13.4	1892	20	AAZ24890	Human secreted pro
36	31.4	13.4	1895	22	AAD08658	Human cancer assoc
37	31.4	13.4	1898	22	AAAF30056	Human cDNA encodin
38	31.4	13.4	1899	20	AAAX06785	Human adult testis
39	31.4	13.4	1899	22	AAAC91465	Human PRO306 cDNA
40	31.4	13.4	1912	21	AAZ50927	Human Protease and
41	31.4	13.4	2061	19	AAV57439	Human protein kina
42	31.4	13.4	2061	21	AAAA47744	HYAK3-2 alpha codi
43	31.4	13.4	2061	22	AAD06042	Human YAK3a (HYAK3
44	31.4	13.4	2140	21	AAZ46161	cDNA sequence enco
45	31.4	13.4	2141	22	AAAF44689	Novel protein kina

#### ALIGNMENTS

RESULT 1  
AAC77198/c  
ID AAC77198 standard; cDNA; 1099 BP.

XX

AC AAC77198;

XX

DT 08-FEB-2001 (first entry)

XX

DE Human ORFX ORF2753 polynucleotide sequence SEQ ID NO:5505.

XX

XX Human; open reading frame; ORFX; detection; hepatotropic;

XX vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;

KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;

KW hypotensive; dermatological; immunosuppressive; antinflammatory;

KW antitumor; antibacterial; antifungal; antihemetic; antithyroid;

KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;

KW neurodegenerative disorder; osteoarthritis; graft vs host disease;

KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;

KW cholesterol ester storage; systemic lupus erythematosus; infection;

KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;

KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;

KW bone damage; cartilage damage; antinflammatory disease; coagulation;

KW thrombosis; contraceptive; ss.

XX Homo sapiens.

OS WO200058473-A2.

XX

PN 05-OCT-2000.

XX

PD 31-MAR-2000; 2000WO-US08621.

XX

XX 31-MAR-1999; 99US-0127607.

XX

```
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX (CURA-) CURAGEN CORP.
XX Shimkets RA, Leach M;
PI
DR WPI: 2000-602362/57.
DR P-PSDB; AAB42989.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 5; Page 4687; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antiprosiatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 1099 BP; 222 A; 339 C; 332 G; 206 T; 0 other;

Query Match 15.0%; Score 35.2; DB 21; Length 1099;
Best Local Similarity 54.7%; Pred. No. 0.19;
Matches 70; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy 67 ctctcatagcaaatactatggttaccagtcctccgaggaaggaggaatccaaacctggaa 126
Db 237 CTCCTCCGAGGCACTTTCGCTTCCAGAGGGTCACGAGGTCGATGAGGTCGCCGACAGGCT 178
Qy 127 gaagacaaggcagtagaagtggcgagtgaggcagcagaagagtggtgacagatcagg 186
Db 177 GCAGACACGAGCGGCAGTTCAGCAGCAGCTGGCGGCGGCAGCTGCGTGACAGCTCCAG 118
Qy 187 gagggtgt 194
Db 117 CAGGATGT 110

RESULT 2
AAV09703/c
ID AAV09703 standard; DNA; 1974 BP.
XX
XX AAV09703;
AC
XX 20-MAY-1998 (first entry)
DT
XX Porcine retrovirus Raji clone ENV DNA.
DE
XX Porcine retrovirus: PoEV; ENV protein; envelope protein; vaccine;
KW diagnosis; xenotransplantation; prophylactic; therapeutic; ds.
XX Porcine retrovirus.
OS
```

```
XX Key Location/Qualifiers
FH 1..1974
FT CDS
FT /*tag= a
FT /product= ENV protein
FT /note= "envelope protein"
XX
XX WO9740167-A1.
XX
XX 30-OCT-1997.
XX
XX 18-APR-1997; 97WO-GB01087.
XX
XX 19-FEB-1997; 97GB-0002668.
XX
XX 10-APR-1996; 96GB-0008164.
XX
XX (IMUT-) IMUTRAN LTD.
XX (OONE-) Q-ONE BIOTECH LTD.
XX Galbraith DN, Haworth C, Lees GM, Smith KT;
XX
XX WPI; 1997-535851/49.
XX P-PSDB; AAW39274.
XX
XX Polynucleotide encoding porcine retrovirus expression product -
PT useful to develop products for use in vaccines, diagnosis and
PT xeno-transplantation
XX
XX Claim 6; Fig 4; 69pp; English.
XX
XX This sequence encodes the porcine retrovirus envelope (ENV) protein
CC isolated from the human cell line Raji. Such viral proteins can be used
CC to develop viral vaccines, antisense nucleic acids, ribozymes and other
CC antiviral agents. They can also be used in xeno-transplantation
CC technology and as diagnostic tools.
XX
SQ Sequence 1974 BP; 534 A; 489 C; 483 G; 468 T; 0 other;

Query Match 14.8%; Score 34.8; DB 18; Length 1974;
Best Local Similarity 55.9%; Pred. No. 0.31;
Matches 66; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 106 gggagagatccaaacctggaaagacgaagcagtagaagtgggcgagggagcgagga 165
Db 986 GAGGATACACAAAGCCACAGACAGAGAGTGCATCAGGTCGGTGAGTTGATGCTTGG 927
Qy 166 aaggttggtgacagatcaggaggggtgctgtgacctttttttgtgaggaattcttagg 223
Db 926 AAAGTCCCTGGATGAGACTGAAGAGTCTCTCTCTCTTAAACAGGAACACCTGGG 869

RESULT 3
AAV82749/c
ID AAV82749 standard; DNA; 3482 BP.
XX
XX AAV82749;
AC
XX 25-FEB-1999 (first entry)
DT
XX Pig endogenous retrovirus (PERV)-B envelope (env) gene region.
DE
XX Pig endogenous retrovirus: PERV-A; envelope protein; PERV-B; subtype;
KW probe; primer; detection; retrovirus; human tissue; xenotransplant;
KW primary porcine tissue; human cell line; porcine cell line; ss.
XX Pig endogenous retrovirus.
OS
XX Key Location/Qualifiers
FH 911..2884
FT CDS
FT /*tag= a
FT /product= envelope_protein
XX
```



```
FT      /*tag= c
FT      /note= "U5 LTR domain"
FT      588..2162
FT      /*tag= d
FT      /product= GAG protein
FT      /note= "viral core polypeptide"
FT      2163..5747
FT      /*tag= e
FT      /product= POL protein
FT      /note= "polymerase peptide"
FT      5620..7590
FT      /*tag= f
FT      /product= ENV protein
FT      /note= "envelope protein"
FT      7591..8209
FT      /*tag= g
FT      7638..8106
FT      /*tag= h
FT      /note= "U3 LTR domain"
FT      7643..7686
FT      /*tag= i
FT      /note= "Binding site for transcription factors
FT      OCT-1, c-Myb, LyF-1, E47 and ETS-1"
FT      7719..7739
FT      /*tag= j
FT      /note= "Binding site for transcription factors
FT      AP-4 and NF1"
FT      7744..7754
FT      /*tag= k
FT      /note= "Binding site for transcription factor AP-1/TR"
FT      7773..7785
FT      /*tag= l
FT      /note= "Binding site for transcription factor ETS-1/GATA"
FT      7793..7834
FT      /*tag= m
FT      /note= "Binding site for transcription factors
FT      GATA, ETS-1, c-Myb, AP-1, GATA, E47"
FT      7841..7851
FT      /*tag= n
FT      /note= "Binding site for transcription factor AP-1"
FT      7873..7883
FT      /*tag= o
FT      /note= "Binding site for transcription factor AP-1"
FT      7885..7914
FT      /*tag= p
FT      /note= "Binding site for transcription factors
FT      AP-1/GATA, c-Myb and AP-1"
FT      7928..7953
FT      /*tag= q
FT      /note= "Binding site for transcription factors E47
FT      and ETS-1"
FT      7964..7968
FT      /*tag= r
FT      7985..7997
FT      /*tag= s
FT      /note= "Binding site for transcription factor GATA"
FT      8003..8007
FT      /*tag= t
FT      8024..8036
FT      /*tag= u
FT      /note= "Binding site for transcription factor GATA"
FT      8042..8046
FT      /*tag= v
FT      8049..8060
FT      /*tag= w
FT      /note= "Binding site for transcription factor
FT      AP-1/CREB"
FT      8075..8090
FT      /*tag= x
FT      1..61
FT      /*tag= y
FT      /note= "R LTR domain"
FT      8127..8135
```

```
FT      /*tag= z
FT      /note= "PolyA downstream element"
FT      8161..8166
FT      /*tag= aa
XX      WO9740167-A1.
PN      30-OCT-1997.
PD      18-APR-1997; 97WO-GB01087.
XX      10-FEB-1997; 97GB-0002668.
PR      19-APR-1996; 96GB-0008164.
XX      (IMUT-) IMUTRAN LTD.
PA      (QONE-) Q-ONE BIOTECH LTD.
XX      Galbraith DN, Haworth C, Lees GM, Smith KT;
PI      WPI; 1997-535851/49.
XX      P-PSDB; AAW39271; AAW39272; AAW39273.
DR      Polynucleotide encoding porcine retrovirus expression product -
XX      useful to develop products for use in vaccines, diagnosis and
PT      xeno-transplantation
PS      Claim 4; Fig 3; 69pp; English.
XX      This DNA sequence encodes the porcine retrovirus (PoEV) virion core
CC      polypeptide (GAG), polymerase (POL) and envelope (ENV) proteins and
CC      also includes the Long Terminal Repeat (LTR). These proteins can be used
CC      to develop viral vaccines, antisense nucleic acids, ribozymes and other
CC      antiviral agents. They can also be used in xeno-transplantation
CC      technology and as diagnostic tools.
XX      Sequence 8209 BP; 2168 A; 2064 C; 2154 G; 1823 T; 0 other;
SQ      Query Match      14.8%; Score 34.8; DB 18; Length 8209;
      Best Local Similarity 55.9%; Pred. No. 0 51;
      Matches 66; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
QY      106 ggggagatccaaacctggagaagaagacagcagtagaagtgggcagtgaggagcagga 165
      ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      6602 GAGGATAGACAAAGCCAAACAAAGAAGTGGCATCAGGTCGGTGGAGTTGATGGCTTGG 6543
QY      166 aaggttggtgacagatcaggaggagggtgtctgacctttttcttgaggaattcttagg 223
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      6542 AAAGCTCCCTGGATGAGACTGAAGAGTCTCTGTCTTCTTAACAGGAACACTGGGG 6485
RESULT 6
AAC93651/c
ID AAC93651 standard; cDNA; 500 BP.
XX
AC AAC93651;
XX
DT 19-FEB-2001 (first entry)
XX
DE Cat flea head and nerve cord (HNC) cDNA, SEQ ID NO:140.
XX
KW Cat flea; head and nerve cord nucleic acid; HNC; flea infestation;
KW vaccine; antiparasitic; therapeutic target; diagnosis; detection; ss.
XX
OS Ctenocephalides felis.
XX
PN WO200061621-A2.
XX
PD 19-OCT-2000.
XX
PF 07-APR-2000; 2000WO-US09437.
XX
PR 09-APR-1999; 99US-0128704.
```





PN WO200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00663.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488897/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human placenta.

XX Claim 25; SEQ ID No 16105; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENP).

XX The present sequence is one such probe. The probes are useful for

XX producing a microarray for predicting, measuring and displaying gene

XX expression in samples derived from human placenta. The probes are useful

XX for antenatal diagnosis of human genetic disorders.

XX SQ Sequence 626 BP; 100 A; 214 C; 104 G; 208 T; 0 other;

Query Match 14.6%; Score 34.4; DB 22; Length 626;

Best Local Similarity 51.3%; Pred. No. 0.27; Indels 0; Gaps 0;

Matches 80; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

Qy 35 tgtccaaggaggttagagagataagagtcaggctctctatagccaaatactatggcttaca 94

Db 384 TGAAGTGGAGGAGGAGGTTGCAGTGGAGCCATATCATGCGCACTATCTACTCCAGCCTGGG 325

Qy 95 gtcccgaggaaggagagatccaaacccctgggaagacaagcagtagaagtggtggcgagt 154

Db 324 TGACAGACGCGAGACTCCATCAAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 265

Qy 155 gggagggcaggaagggttggtgacagatcaggagg 190

Db 264 AAGAAGGGGAAGGGGAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 229

RESULT 9

AAI07823/c

ID AAI07823 standard; DNA; 626 BP.

XX AAI07823;

XX 09-OCT-2001 (first entry)

XX Probe #7814 used to measure gene expression in human breast sample.

XX Probe; human; breast disease; breast cancer; development disorder; ss;

XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.

XX Homo sapiens.

XX WO200157270-A2.

XX 09-AUG-2001.

XX 29-JAN-2001; 2001WO-US00661.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-476286/51.

XX Novel single exon nucleic acid probe used to measuring gene expression

XX in a human breast.

XX Claim 25; SEQ ID No 7814; 322pp; English.

XX The present invention relates to novel single exon nucleic acid probes.

XX The present sequence is one such probe. The probes are useful for

XX measuring human gene expression in a human breast sample, where the probe

XX hybridises at high stringency to a nucleic acid expressed in the human

XX breast. The probes are useful for predicting, diagnosing, grading,

XX staging, monitoring and prognosing diseases of the human breast,

XX particularly those diseases with polygenic aetiology. The diseases

XX include: breast cancer, disorders of development, inflammatory diseases

XX of the breast, fibrocystic changes, proliferative breast disease and

XX non-carcinoma tumours.

XX Note: The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 626 BP; 100 A; 214 C; 104 G; 208 T; 0 other;

Query Match 14.6%; Score 34.4; DB 22; Length 626;

Best Local Similarity 51.3%; Pred. No. 0.27; Indels 0; Gaps 0;

Matches 80; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

Qy 35 tgtccaaggaggttagagagataagagtcaggctctctatagccaaatactatggcttaca 94

Db 384 TGAAGTGGAGGAGGAGGTTGCAGTGGAGCCATATCATGCGCACTATCTACTCCAGCCTGGG 325

Qy 95 gtcccgaggaaggagagatccaaacccctgggaagacaagcagtagaagtggtggcgagt 154

Db 324 TGACAGACGCGAGACTCCATCAAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 265

Qy 155 gggagggcaggaagggttggtgacagatcaggagg 190

Db 264 AAGAAGGGGAAGGGGAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 229

RESULT 10

AAT65102/c

ID AAT65102 standard; DNA; 2784 BP.

XX AAT65102;

XX 24-FEB-1998 (first entry)

XX Luciferase gene under control of rat MLC-2 gene promoter.

XX Luciferase; upstream regulatory region; MLC-2 promoter; gene therapy;

XX rat cardiac myosin light chain gene; heart disease; cardiac muscle; ss.

XX Chimeric - Synthetic.

XX Chimeric - Rattus sp.

XX Chimeric - Adenovirus.

XX Key Location/Qualifiers

FT misc\_feature 1..156

FT /\*tag= a

FT /note= "encodes Adenovirus Ad5 psi encapsidation"



CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.

XX Sequence 1954 BP; 439 A; 504 C; 537 G; 474 T; 0 other;

Query Match 14.5%; Score 34; DB 22; Length 1954;  
 Best Local Similarity 59.2%; Pred. No. 0.54;  
 Matches 58; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 104 aaggggagatccaaaccctggaagaagacagcagtagaagtggcgagtgaggagcag 163

Db 194 aaggggagattcagcccggaacctggcgatggcgtggaagcgtggaacaggaaggaa 253

Qy 164 gaaaggttggtgacagatcagggaggggtgtctgacct 201

Db 254 aggcggtctgggagagagcgcgagctggtgtgtact 291

# RESULT 12

AAI59351  
 ID AAI59351 standard; cDNA; 2182 BP.

AC AAI59351;

DT -22-OCT-2001 (first entry)

XX Human polynucleotide SEQ ID NO 1554.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia; ss.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

XX P-PSDB; AAM40195.

XX Novel nucleic acids and polypeptides, useful for treating disorders

XX such as central nervous system injuries -

XX Claim 1; SEQ ID NO 1554; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and

XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,

XX immunosuppressant and cytostatic activity. The polynucleotides are useful

XX

CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.

XX Sequence 2182 BP; 477 A; 552 C; 587 G; 566 T; 0 other;

Query Match 14.5%; Score 34; DB 22; Length 2182;

Best Local Similarity 59.2%; Pred. No. 0.57;

Matches 58; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 104 aagggagatccaaaccctggaagaagacagcagtagaagtggcgagtgaggagcag 163

Db 429 aaggggagattcagcccggaacctggcgatggcgtggaagcgtggaacaggaaggaa 488

Qy 164 gaaaggttggtgacagatcagggaggggtgtctgacct 201

Db 489 agcggtctgggagagagcgcgagctggtgtgtact 526

# RESULT 13

AAAD06295

ID AAD06295 standard; cDNA; 2184 BP.

XX AC AAD06295;

XX 10-AUG-2001 (first entry)

XX Human Ras suppressor sur-5 cDNA.

XX Ras suppressor; sur-5; cytostatic; Ras signalling; gene therapy;  
 KW cell growth; differentiation; proliferation; transgenic animal;  
 KW cancer; human; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 1..1998

XX FT /\*tag= a

XX FT /product= "SUR-5 protein"

XX FT /transl\_except= (pos:1756..1758, aa:Asp)

XX FT /note= "CDS does not include start codon"

XX FT /partial

XX US6225456-B1.

XX 01-MAY-2001.

XX 06-MAY-1999; 99US-0307265.

XX 07-MAY-1998; 98US-0084590.

XX (UYTE-) UNIV TECHNOLOGY CORP.

XX Gu T, Orita S, Han M;

XX WPI; 2001-342417/36.

XX P-PSDB; AAE02345.

XX Novel Ras suppressor SUR-5 polynucleotides from human, murine,

XX Caenorhabditis elegans, for treating Ras-mediated effects, especially

XX cancer and producing SUR-5 polypeptides for screening drugs altering

XX Ras signalling -



Search completed: January 8, 2002, 08:52:12  
Job time: 5834 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: January 8, 2002, 08:05:20 ; Search time 2296.3 Seconds  
(without alignments)  
1688.297 Million cell updates/sec

Title: US-09-823-101-2  
Perfect score: 235  
Sequence: 1 acctctaccactgctgac.....ttcttaggcaagtgaagctt 235

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_om:\*  
20: em\_or:\*  
21: em\_ov:\*  
22: em\_pat:\*  
23: em\_ph:\*  
24: em\_pl:\*  
25: em\_ro:\*  
26: em\_sts:\*  
27: em\_sy:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htgo\_hum:\*  
31: em\_htgo\_inv:\*  
32: em\_htgo\_rod:\*  
33: em\_htg\_hum:\*  
34: em\_htg\_inv:\*  
35: em\_htg\_rod:\*  
36: em\_htg\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	43.8	18.6	7218	6	I66494	166494 Sequence 14
C 2	39.6	16.9	176697	2	AC021170	AC021170 Homo sapi
C 3	38.4	16.3	161261	9	CNS05TE3	AL358332 Human chr
C 4	38	16.2	150869	2	AC023992	AC023992 Homo sapi
C 5	38	16.2	158905	9	AC005666	AC005666 Homo sapi
C 6	37.8	16.1	123149	9	AC008655	AC008655 Homo sapi
C 7	37.8	16.1	198316	2	AC019157	AC019157 Homo sapi
C 8	37.4	15.9	153773	9	HUAC004020	AC004020 Homo sapi
C 9	37	15.7	161081	2	AC092015	AC092015 Homo sapi
C 10	37	15.7	161119	33	AL356368	AL356368 Human DNA
C 11	37	15.7	162518	2	AL592288	AL592288 Homo sapi
C 12	36.8	15.7	124288	2	AF207067	AF207067 Homo sapi
C 13	36.8	15.7	156833	9	AC018712	AC018712 Homo sapi
C 14	36.8	15.7	180192	2	AC020681	AC020681 Homo sapi
C 15	36.6	15.6	263	11	HUMUT5082A	L31021 Human STS U
C 16	36.6	15.6	889	5	AF089018	AF089018 Cacicus s
C 17	36.6	15.6	889	5	AF290169	AF290169 Cacicus s
C 18	36.6	15.6	77649	2	AC016654	AC016654 Homo sapi
C 19	36.6	15.6	112901	2	AP002511	AP002511 Homo sapi
C 20	36.6	15.6	152472	9	AP000486	AP000486 Homo sapi
C 21	36.6	15.6	161740	2	AC084778	AC084778 Homo sapi
C 22	36.6	15.6	164577	2	AC016549	AC016549 Homo sapi
C 23	36.6	15.6	171411	2	AC087609	AC087609 Homo sapi
C 24	36.6	15.6	177573	2	AC019076	AC019076 Homo sapi
C 25	36.6	15.6	177638	9	AC016652	AC016652 Homo sapi
C 26	36.6	15.6	178105	9	AC073057	AC073057 Homo sapi
C 27	36.6	15.6	203876	2	AC010413	AC010413 Homo sapi
C 28	36.2	15.4	141385	2	AC090714	AC090714 Oryza sat
C 29	36.2	15.4	170431	2	AC093331	AC093331 Homo sapi
C 30	36.2	15.4	172813	2	AC015720	AC015720 Homo sapi
C 31	36.2	15.4	174305	2	AC012191	AC012191 Homo sapi
C 32	36.2	15.4	196416	9	AC006483	AC006483 Homo sapi
C 33	36.2	15.4	251993	2	AC023822	AC023822 Homo sapi
C 34	36	15.3	2805	4	AF286474	AF286474 Bos tauru
C 35	36	15.3	117071	9	AL133375	AL133375 Human DNA
C 36	36	15.3	130526	9	AL133282	AL133282 Human DNA
C 37	36	15.3	163154	2	AC026532	AC026532 Homo sapi
C 38	36	15.3	166942	9	HS1054A22	AL031651 Human DNA
C 39	36	15.3	194405	2	AC032004	AC032004 Homo sapi
C 40	35.8	15.2	1968	14	PEN288589	AJ288589 Porcine e
C 41	35.8	15.2	158608	2	AC026834	AC026834 Homo sapi
C 42	35.8	15.2	220399	2	AC015918	AC015918 Homo sapi
C 43	35.6	15.1	115915	2	AP000859	AP000859 Homo sapi
C 44	35.6	15.1	153588	9	AC010533	AC010533 Homo sapi
C 45	35.6	15.1	161797	2	AP001775	AP001775 Homo sapi

ALIGNMENTS

RESULT 1  
LOCUS I66494/C  
DEFINITION Sequence 14 from patent US 5670367.  
ACCESSION I66494  
VERSION I66494.1 GI:2724471  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 7218)  
AUTHORS Dörner, F., Scheiflinger, F. and Falkner, F. Gunter.  
TITLE Recombinant fowlpox virus  
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;  
FEATURES  
source Location/Qualifiers  
1..7218  
/organism="unknown"  
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others  
ORIGIN

28-DEC-1997









repeat_region	/rpt_family="L1MB3" 2471. .2778
repeat_region	/rpt_family="AluY" 2779. .2925
repeat_region	/rpt_family="AluJo/FRAM" 2926. .2948
repeat_region	/rpt_family="(CAAA)n" 2955. .3255
repeat_region	/rpt_family="L1MB3" 3256. .3560
repeat_region	/rpt_family="AluSc" 3570. .3729
repeat_region	/rpt_family="FAM" 3730. .3787
repeat_region	/rpt_family="L1MB3" complement(3964. .4114)
repeat_region	/rpt_family="MER5B" 5111. .5426
repeat_region	/rpt_family="AluJb" 5480. .5515
repeat_region	/rpt_family="AT_rich" 5752. .5784
repeat_region	/rpt_family="AT_rich" complement(5999. .6328)
repeat_region	/rpt_family="MER58B" 6435. .6518
repeat_region	/rpt_family="L1MB5" 6519. .6815
repeat_region	/rpt_family="AluY" 6816. .7533
repeat_region	/rpt_family="L1MB5" 7535. .7694
repeat_region	/rpt_family="AluJb" 7705. .8001
repeat_region	/rpt_family="AluSq" 8124. .8183
repeat_region	/rpt_family="MLT2D" complement(8184. .8205)
repeat_region	/rpt_family="AluSq" complement(8206. .8229)
repeat_region	/rpt_family="(CAAAA)n" complement(8230. .8515)
repeat_region	/rpt_family="AluSq" 8516. .8942
repeat_region	/rpt_family="MLT2D" complement(8957. .9079)
repeat_region	/rpt_family="MER7A" 9080. .9419
repeat_region	/rpt_family="THEIC" complement(9420. .9626)
repeat_region	/rpt_family="MER7A" 9702. .9748
repeat_region	/rpt_family="MLT2D" 9863. .9940
repeat_region	/rpt_family="L2" 10022. .10238
repeat_region	/rpt_family="L1MB3" complement(10325. .11142)
repeat_region	/rpt_family="L2" complement(11143. .11173)
repeat_region	/rpt_family="(TGAA)n" complement(11175. .11440)
repeat_region	/rpt_family="AluJb" complement(11441. .12055)
repeat_region	/rpt_family="L2" complement(12056. .12364)
repeat_region	/rpt_family="AluSq" complement(12365. .12447)
repeat_region	/rpt_family="L2" 12448. .12722
repeat_region	/rpt_family="AluSc" 12726. .13024
repeat_region	/rpt_family="AluJb"

repeat_region	complement(13025..13417)
repeat_region	/rpt_family="L2"
repeat_region	13419..13698
repeat_region	/rpt_family="AluSx"
repeat_region	13699..13726
repeat_region	/rpt_family="(TAAA)n"
repeat_region	13783..14076
repeat_region	/rpt_family="AluY"
repeat_region	complement(14083..14322)
repeat_region	/rpt_family="MIR"
repeat_region	14484..14510
repeat_region	/rpt_family="AT-rich"
repeat_region	complement(15900..16212)
repeat_region	/rpt_family="L1MA4A"
repeat_region	16280..16596
repeat_region	/rpt_family="MER7A"
repeat_region	16986..17306
repeat_region	/rpt_family="AluJo"
repeat_region	17422..17468
repeat_region	/rpt_family="(CAAAA)n"
repeat_region	17550..17642
repeat_region	/rpt_family="HAL1"
repeat_region	complement(17643..17937)
repeat_region	/rpt_family="AluSp"
repeat_region	17938..18133
repeat_region	/rpt_family="HAL1"
repeat_region	18764..19102
repeat_region	/rpt_family="L1MB"
repeat_region	19121..19742
repeat_region	/rpt_family="L1M4"
repeat_region	20318..20493
repeat_region	/rpt_family="MLT1J"
repeat_region	complement(20498..20831)
repeat_region	/rpt_family="L1M4"
repeat_region	complement(20888..21059)
repeat_region	/rpt_family="AluSg"
repeat_region	complement(21060..21328)
repeat_region	/rpt_family="AluY"
repeat_region	complement(21329..21460)
repeat_region	/rpt_family="AluSg"
repeat_region	22123..23454
repeat_region	/rpt_family="L1MB2"
repeat_region	23455..23586
repeat_region	/rpt_family="FLAM_C"
repeat_region	23587..24009
repeat_region	/rpt_family="L1MB2"
repeat_region	24010..24312
repeat_region	/rpt_family="AluSg"
repeat_region	24333..25000
repeat_region	/rpt_family="L1M1"
repeat_region	25001..25298
repeat_region	/rpt_family="AluSg"
repeat_region	25299..26025
repeat_region	/rpt_family="L1M1"
repeat_region	26026..26321
repeat_region	/rpt_family="AluSx"
repeat_region	26335..27573
repeat_region	/rpt_family="L1M1"
repeat_region	27574..27870

Query Match  
16.2%; Score 38; DB 9; Length 158905;

Best Local Similarity 62.88; Pred. No. 1.3;

Best local similarity	02.00;	REC. NO: 1.5;
Matches	59; Conservative	0; Mismatches
		35; Indels
		0; Gaps
		0;

Ov 124 aaagaaacaaaggcagtagaaagtggccgagtggaagaagqattggaqatc 183

[illegible]

Ov 184 agggagggtgtctgacctttttcttgaaggaatt 217

[illegible]

```

RESULT 6
AC008655/c
LOCUS      AC008655      123149 bp      DNA      PRI      15-AUG-2001
DEFINITION Homo sapiens chromosome 19 clone CTB-191K22, complete sequence.
ACCESSION  AC008655
VERSION    AC008655.7      GI:15187196
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 123149)
AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE     Direct Submission
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 123149)
AUTHORS   DOE Joint Genome Institute.
TITLE     Direct Submission
JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE  3 (bases 1 to 123149)
AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE     Direct Submission
JOURNAL   Submitted (28-FEB-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE  4 (bases 1 to 123149)
AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE     Direct Submission
JOURNAL   Submitted (15-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT   On Aug 15, 2001 this sequence version replaced gi:13162496.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www.sngc.stanford.edu
            Quality: Phrap Quality >=40 99.8% of Sequence;
            Estimated Total Number of Errors is 0.2.

FEATURES             source
     source          1..123149
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="CTB-191K22"
BASE COUNT  29370 a 30796 c 33371 g 29612 t
ORIGIN

Query Match      16.1%; Score 37.8; DB 9; Length 123149;
Best Local Similarity 54.7%; Pred. No. 1.4;
Matches 75; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy  26  gagtgggggtgtccaaaggagtagagagataaagtcaggctctcatagccaaatacta 85
      || ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  51202 GACGGGGTTAGAGGGGTGATGGGGGTCTCTGAAAAGGAACGGAGACCCAGAGAGAG 51143

Qy  86  tggcttacgtcccgaggagagagatcccaacccttgaaagacagcagtagaag 145
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  51142 GGGGNACAGGTTCGCCAGGAAGTGGAGACAGACAGACCTAGAAAAGAGAGACCCAGGAAG 51083

Qy  146  tgggcgagtgaggagca 162
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  51082 GTGAACATAGGAGGAA 51066

RESULT 7
AC019157
LOCUS      AC019157      198316 bp      DNA      HTG      07-JUL-2000
DEFINITION Homo sapiens chromosome 19 clone RP11-25A12, WORKING DRAFT
            SEQUENCE, 35 unordered pieces.
ACCESSION  AC019157
VERSION    AC019157.5      GI:8569727
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     human.

```

```

ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 198316)
AUTHORS   Waterston,R.H.
TITLE     The sequence of Homo sapiens clone
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 198316)
AUTHORS   Waterston,R.H.
TITLE     Direct Submission
JOURNAL   Submitted (30-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT   On Jun 16, 2000 this sequence version replaced gi:7230833.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0025A12
----- Summary Statistics -----
Sequencing vector: M13; 88%
Sequencing vector: plasmid; 12%
Chemistry: Dye-primers; 88% of reads
Chemistry: Dye-terminator Big Dye; 12% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 175030 bases at least Q40
Consensus quality: 182010 bases at least Q30
Consensus quality: 185613 bases at least Q20
Insert size: 233000; agarose-fp
Insert size: 194916; sum-of-contigs
Quality coverage: 3.45 in Q20 bases; agarose-fp
Quality coverage: 4.12 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 35 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1530: contig of 1530 bp in length
* 1531: gap of unknown length
* 1630: contig of 1530 bp in length
* 1631: gap of 1155 bp in length
* 2785: contig of 1155 bp in length
* 2885: gap of unknown length
* 4726: contig of 1841 bp in length
* 4826: gap of unknown length
* 4727: gap of unknown length
* 6177: contig of 1351 bp in length
* 6178: gap of unknown length
* 6278: contig of 1331 bp in length
* 7608: gap of unknown length
* 7709: contig of 2349 bp in length
* 10058: gap of unknown length
* 10158: contig of 1419 bp in length
* 11577: gap of unknown length
* 11677: contig of 1175 bp in length
* 12851: gap of unknown length
* 12952: contig of 1517 bp in length
* 14468: gap of unknown length
* 14569: gap of unknown length
* 14569: contig of 2242 bp in length
* 16810: gap of unknown length
* 16811: contig of 1791 bp in length
* 18701: gap of unknown length
* 18801: contig of 2442 bp in length
* 21243: gap of unknown length
* 21244: contig of 3434 bp in length
* 21344: gap of unknown length
* 24778: contig of 3723 bp in length
* 24878: gap of unknown length
* 28601: contig of 3939 bp in length
* 28701: gap of unknown length
* 32639: contig of 3939 bp in length
* 32640: gap of unknown length

```











Siddiqui, R., Taudien, S., Wen, G., Korenberg, J.R., Rosenthal, A. and Platzer, M.

## TITLE Chromosome 8 genomic sequence

Unpublished  
2 (bases 1 to 124288)

REFERENCE Reichwald, K., Wen, G., Schillhabel, M., Baumgart, C., Dette, M., Menzel, U., Jahn, N. and Rosenthal, A.

## TITLE Direct Submission

JOURNAL Submitted (19-NOV-1999) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany  
On Jun 2, 2001 this sequence version replaced gi:8152098.

## COMMENT

----- Genome Center  
Center: Institute of Molecular Biotechnology

Center code: IMB

Web site: <http://genome.imb-jena.de/>

Contact: [gscj-submit@genome.imb-jena.de](mailto:gscj-submit@genome.imb-jena.de)

----- Project Information

Center project name: H208

Center clone name: XX-CTB875C8

----- Summary Statistics

Sequencing vector: M13; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 117740 bases at least Q40

Consensus quality: 120559 bases at least Q30

Consensus quality: 123333 bases at least Q20

Quality coverage: 5.69 x in Q20 bases; sum-of-contigs

-----

## Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

-----

\* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 4113: contig of 4113 bp in length

\* 4114 4213: gap of unknown length

\* 4214 47267: contig of 43054 bp in length

\* 47268 47367: gap of unknown length

\* 47368 124288: contig of 76921 bp in length.

Location/Qualifiers

1. .124288

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="8"

/map="Bq24.1"

/clone="XX-CTB875C8"

1. .4

/note="assembly\_fragment"

clone\_end:SP6

vector\_side:left

124285. .124288

/note="assembly\_fragment"

clone\_end:T7

vector\_side:right

BASE COUNT 36681 a 26600 c 25566 g 35241 t 200 others

## ORIGIN

-----

Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
----- Summary Statistics  
-----  
Center project name: H\_NH0343H10

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

QY 100 gaggaaggagatccaaacccctgcaaaagacaaagcagtagaagtgccagtgaggag 159  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 14865 GAGGAGGAGGAGGAGACATGGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAG 14806  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 160 gcagaaagggttggtgcagatcagggagggtgtctgaccttttcttgaggaaattct 219  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 14805 GCAGAAAGGAGGAGAGACAGACAGAAAGAGTCTACCATTTCTCGAGCTGATCCGATGTAT 14746  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 220 taggcaag 227  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 14745 TAGGTAAG 14738  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

## RESULT 13

AC018712/c

LOCUS AC018712 156833 bp DNA PRI 09-MAY-2001

DEFINITION Homo sapiens BAC clone RP11-343H10 from 2, complete sequence.

ACCESSION AC018712

VERSION AC018712.5 GI:11415210

KEYWORDS HTG.

SOURCE human.

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 156833)

TITLE Toward a complete human genome sequence

JOURNAL Genome Res. 8 (11), 1097-1108 (1998)

MEDLINE 99063792

REFERENCE 2 (bases 1 to 156833)

AUTHORS Edwards, J., Maupin, R., Hawkins, M. and Smith, R.

TITLE The sequence of Homo sapiens BAC clone RP11-343H10

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 156833),

AUTHORS Waterston, R.H.

TITLE Direct Submission

JOURNAL Submitted (16-DEC-1999) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 4 (bases 1 to 156833)

AUTHORS Waterston, R.H.

TITLE Direct Submission

JOURNAL Submitted (29-NOV-2000) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 5 (bases 1 to 156833)

AUTHORS Waterston, R.

TITLE Direct Submission

JOURNAL Submitted (09-MAY-2001) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

COMMENT On Nov 29, 2000 this sequence version replaced gi:9838147.

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>

Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)

----- Summary Statistics

-----  
Center project name: H\_NH0343H10

## misc\_feature

clone\_end:SP6

vector\_side:left

124285. .124288

/note="assembly\_fragment"

clone\_end:T7

vector\_side:right

BASE COUNT 36681 a 26600 c 25566 g 35241 t 200 others

ORIGIN

-----

Query Match 15.7%; Score 36.8; DB 2; Length 124288;

Best Local Similarity 55.5%; Pred. No. 2.8;

Matches 71; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

(VCTOR: PBAC3.6)

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-313011, 200 bp overlap; the clone sequenced to the right is AC020681. Actual start of this clone is at base position 198520 of RP11-313011; actual end is at base position 156833 of RP11-343H10.

Bases 123072 to 123647 of RP11-343H10 consists only of a PCR product from clone DNA. Bases 62320 to 62338 consists only of a single M13 subclone. Digest and PCR information confirms the assembly.

## FEATURES

.source	Location/Qualifiers
	1..156833
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-343H10"
	/clone_lib="RPCI-11"
repeat_region	937..1272
	/rpt_family="L2"
repeat_region	1280..1408
	/rpt_family="L2"
repeat_region	1639..2020
	/rpt_family="L1"
repeat_region	2021..2389
	/rpt_family="MaLR"
repeat_region	3971..4333
	/rpt_family="MaLR"
repeat_region	4334..4849
	/rpt_family="L1"
repeat_region	4850..5138
	/rpt_family="Alu"
repeat_region	5139..5421
	/rpt_family="L1"
repeat_region	5700..6000
	/rpt_family="Alu"
repeat_region	6914..8085
	/rpt_family="L1"
repeat_region	8086..8143
	/rpt_family="MIR"
repeat_region	8172..8324
	/rpt_family="MIR"
repeat_region	8798..8978
	/rpt_family="MaLR"
repeat_region	10040..10639
	/rpt_family="L1"
repeat_region	10640..10945
	/rpt_family="Alu"
repeat_region	10946..10995
	/rpt_family="L1"
repeat_region	11259..11317
	/rpt_family="L2"
repeat_region	12438..12762
	/rpt_family="MERL_type"
repeat_region	12942..13248
	/rpt_family="Alu"
repeat_region	14289..14321
	/rpt_family="L2"
repeat_region	18392..18494
	/rpt_family="MIR"
repeat_region	19330..19607
	/rpt_family="L1"
repeat_region	19668..19778
	/rpt_family="L1"
repeat_region	19794..19930
	/rpt_family="Alu"
repeat_region	20395..20617
	/rpt_family="MIR"
misc_feature	21510..21934
	/note="similar to EST AV695578 (NID:g10297441)"
misc_feature	21510..22135
	/note="similar to EST AV684851 (NID:g10286714)"
misc_feature	21510..22057
	/note="similar to EST AV691567 (NID:g10293430)"
repeat_region	21861..22339
	/rpt_family="ERV1"
misc_feature	22106..22228
	/note="similar to EST AV691567 (NID:g10293430)"
repeat_region	22968..23304
	/rpt_family="ERV1"
repeat_region	23305..23608
	/rpt_family="Alu"
repeat_region	23609..23740
	/rpt_family="ERV1"
repeat_region	25588..25870
	/rpt_family="L2"
repeat_region	26252..26701
	/rpt_family="MaLR"
repeat_region	26774..27034
	/rpt_family="L2"
repeat_region	27040..27583
	/rpt_family="L1"
repeat_region	27593..27801
	/rpt_family="L1"
repeat_region	27836..28064
	/rpt_family="MaLR"
repeat_region	28403..28493
	/rpt_family="L1"
repeat_region	28654..28768
	/rpt_family="L1"
repeat_region	28909..29145
	/rpt_family="L1"
repeat_region	29939..30091
	/rpt_family="MERL_type"
misc_feature	30032..30522
	/note="similar to EST AL120277 (NID:g5926176)"
repeat_region	30271..30730
	/rpt_family="MaLR"
repeat_region	30975..31269
	/rpt_family="Alu"
repeat_region	31281..31769
	/rpt_family="L1"
repeat_region	31880..32055
	/rpt_family="L1"
repeat_region	32089..32193
	/rpt_family="L1"
repeat_region	32198..32495
	/rpt_family="Alu"
repeat_region	32534..32943
	/rpt_family="L1"
repeat_region	33070..33734
	/rpt_family="L1"
repeat_region	34622..34704
	/rpt_family="MERL_type"
repeat_region	34700..34860
	/rpt_family="MERL_type"

Query Match

Best Local Similarity 58.0%; Score 36.8; DB 9; Length 156833;

Pred. No. 2.8;

```

Matches 65; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
Qy 110 agatccaaacctgcgaagaagacagcagtagaagtgcagtgaggagcaggaagc 169
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 51633 AGAGGGAGCCTGCAAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 51574
Qy 170 ttgggtgacatcagcagggaggtgtctgacctttttcttgaggaaattctta 221
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 51573 AGAGGGAGGAGAGAGGAGGAGGAGGAGCTTTTCTTAATTAATTTTA 51522

RESULT 14
AC020681/c
LOCUS
DEFINITION
AC020681 180192 bp DNA HTG 19-JUL-2000
Homo sapiens chromosome 2 clone RP11-366L15 map 2, WORKING DRAFT
SEQUENCE, 13 unordered pieces.
ACCESSION
AC020681
VERSION
KEYWORDS
SOURCE
HTG; HTGS_PHASE1; HTGS_DRAFT.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 180192)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 2, clone RP11-366L15
Unpublished
REFERENCE
2 (bases 1 to 180192)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howard,J., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczkyl,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,
Pierre,N., Pisan,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (08-Jan-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 19, 2000 this sequence version replaced gi:6731248.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIPR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3778
Center clone name: 366_L_15
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 173966 bases at least Q40
Consensus quality: 176894 bases at least Q30
Consensus quality: 178103 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 178992; sum-of-contigs
Quality coverage: 5.1 in Q20 bases; agarose-fp
Quality coverage: 5.2 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently

```

\* consists of 13 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1617: contig of 1617 bp in length  
 1618 1717: gap of 100 bp  
 1718 6392: contig of 4675 bp in length  
 6393 6492: gap of 100 bp  
 6493 11804: contig of 5312 bp in length  
 11805 11904: gap of 100 bp  
 11905 18137: contig of 6233 bp in length  
 18138 18237: gap of 100 bp  
 18238 25025: contig of 6788 bp in length  
 25026 25125: gap of 100 bp  
 25126 33095: contig of 7970 bp in length  
 33096 33195: gap of 100 bp  
 33196 46602: contig of 13407 bp in length  
 46603 46702: gap of 100 bp  
 46703 61609: contig of 14907 bp in length  
 61610 61709: gap of 100 bp  
 61710 80467: contig of 18758 bp in length  
 80468 80567: gap of 100 bp  
 80568 97584: contig of 17017 bp in length  
 97585 97684: gap of 100 bp  
 97685 119038: contig of 21354 bp in length  
 119039 119138: gap of 100 bp  
 119139 148789: contig of 29651 bp in length  
 148790 148889: gap of 100 bp  
 148890 180192: contig of 31303 bp in length.

## FEATURES

## source

1..180192  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /map="2"

/clone="RP11-366L15"  
 /clone\_lib="RPC1-11 Human Male BAC"  
 1..1617  
 /note="assembly\_fragment"

## misc\_feature

1718..6392

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

6493..11804

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

11905..18137

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

18238..25025

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

25126..33095

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

61710..80467

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

80568..97584

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

97685..119038

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

119139..148789

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

148890..180192

## misc\_feature

/note="assembly\_fragment"

BASE COUNT 53893 a 36593 c 36153 g 52350 t 1203 others  
 ORIGIN





**THIS PAGE BLANK (USPTO)**